

文章编号: 1000-7032(2026)01-0022-11

无机晶体材料中稀土离子局域配位键长的第一性原理 计算与机器学习预测

陈炳荣^{1†}, 贺凤婷^{1†}, 谭佩雯¹, 谢梦圆¹, 周海琼¹, 尹 浩¹,
李 真¹, 陈振强¹, 马凤凯^{1*}, 苏良碧^{2*}

(1. 暨南大学 光电工程系, 广东省晶体材料和晶体激光技术与应用工程技术研究中心, 广东 广州 510632;

2. 中国科学院上海硅酸盐研究所 功能晶体与器件全国重点实验室, 上海 201899)

摘要: 在无机晶体材料中, 稀土离子所处局域配位环境对其受激辐射性能有显著影响。然而, 快速批量获取稀土离子局域配位结构信息仍充满挑战。本文结合第一性原理计算与机器学习, 实现了无机晶体材料中稀土离子局域结构键长的准确预测。基于训练模型预测了 96 种未报道的稀土掺杂无机晶体材料的局域键长, 并通过第一性原理计算验证了其准确性。该方法简单、快速, 能够批量预测稀土离子局域结构信息, 为稀土掺杂无机晶体材料的结构和性能设计提供了重要途径。

关 键 词: 键长; 局域结构; 稀土掺杂无机晶体

中图分类号: O482.31

文献标识码: A

DOI: 10.37188/CJL.20250215

CSTR: 32170.14.CJL.20250215

A Combined DFT and Machine Learning to Predict Bond Lengths in Rare Earth-doped Inorganic Crystals

CHEN Bingrong^{1†}, HE Fengting^{1†}, TAM YU Puy Mang¹, XIE Mengyuan¹, ZHOU Haiqiong¹,
YIN Hao¹, LI Zhen¹, CHEN Zhenqiang¹, MA Fengkai^{1*}, SU Liangbi^{2*}

(1. Guangdong Provincial Engineering Research Center of Crystal and Laser Technology, Department of Optoelectronic Engineering,
Jinan University, Guangzhou 510632, China;

2. State Key Laboratory of Functional Crystals and Devices, Shanghai Institute of Ceramics, Chinese Academy of Sciences,
Shanghai 201899, China)

* Corresponding Authors, E-mail: mafengkai@jnu.edu.cn; suliangbi@mail.sic.ac.cn

Abstract: Local structures significantly influence the stimulated radiation of rare earth ions in inorganic materials. However, quickly obtaining local structure information remains a challenge. In this investigation, we combined first-principles calculation with machine learning. From the calculation results, it can be shown that the ionic radii, coordination numbers, ligands, and polyhedral shapes of rare earth ions impact the local structural bond lengths of rare earths. Using the BPNN (Back-propagation neural network) and SVR (Support vector regression) algorithms, considering 22 chemical and structural influencing factors, we were able to accurately predict the bond lengths of rare earth ions. The contributions of these chemical features were quantitatively analyzed and found multiple factors affecting the bond lengths. Based on the learned model, the bond lengths of 96 unknown rare earth inorganic crystals were predicted, and validated with first principles calculation. The high feasibility demonstrates that this easy and quick

收稿日期: 2025-09-18; 修订日期: 2025-10-09

基金项目: 国家自然科学基金(62475104); 中国科学院稳定支持基础研究青年团队(YSBR-024); 暨南大学中央高校基础研究业务费(21624407); 广东省稀土开发及应用研究重点实验室开放基金资助项目(XTKY-202402)

Supported by National Natural Science Foundation of China(62475104); Chinese Academy of Sciences (CAS) Stable Support Project for Basic Research Youth Teams (YSBR-024); Jinan University Fundamental Research Funds for the Central Universities (21624407); Open Fund of the Guangdong Provincial Key Laboratory of Rare Earth Development and Application(XTKY-202402)

†: 共同贡献作者

evaluation method for local structures of rare earth ions provides a valuable approach for predicting, designing, and optimizing the local structures and properties of functional inorganic materials.

Keywords: bond lengths; local structures; rare earth-doped inorganic crystal

1 引言

作为基础光源之一,激光广泛应用于工业、国防、医疗等领域,而稀土掺杂的激光晶体材料因其优异的光学、力学、热学性能是产生激光最重要的一类增益介质,如Nd:YAG、Yb:YAG、Nd:YVO₄、Nd:YLF、Tm:YLF等。不同的激光晶体材料产生的激光特性不同,其应用领域不同。例如,Nd:YAG适用于高功率激光^[1],而Nd:YVO₄则适用于中低功率激光^[2]。晶体的性能,特别是光谱性能,受基质晶体影响显著^[3]。激活离子的局域结构决定了其辐射特性,进而影响晶体的激光性能。激活离子局域配位环境被视为激光晶体材料光谱性能的功能基元。因此,研究稀土离子的局域配位结构至关重要,将极大促进晶体结构与性能的关联模型构建,以及新型激光晶体材料的探索和发展。

局域配位结构包括阴离子配体、对称性、键长、多面体形状等信息,而解析该微结构信息的技术手段众多,如X射线吸收光谱(XAS)、电子自旋共振(EPR)、高角环形暗场像(STEM-HAADF)、核磁共振(NMR)、模拟计算等。利用相关技术,人们在解析局域结构方面取得了显著进展。例如,Popov等利用XAS技术研究了钨酸镧 $Ln_2(WO_4)_3$ ($Ln=La, Pr, Nd, Sm, Eu, Gd, Dy$)晶体中稀土离子的局域配位结构,发现随离子半径从La到Dy降低, Ln^{3+} 离子在 $Ln_2(WO_4)_3$ 中的局域环境无序度逐渐增加,同时 $Ln-O$ 键长不断减小^[4]。此外,他们还利用扩展X射线吸收精细结构(EXAFS)解析了 $Ln_2M_2O_7$ ($Ln=La\sim Yb, Y; M=Ti, Zr, Ce$)晶体中稀土离子的键长和配位数的演变规律,与其他表征结果一致^[5]。Krieke等使用EPR观察到两种与O²⁻中心相关的空位缺陷^[6]。STEM-HAADF被用于直接观察Ce³⁺掺杂CaYAlO₄晶体中Ce离子相关的团簇^[7]。Roya等结合NMR和基于密度泛函模拟计算,揭示了 $RE-PO_4$ ($RE=La, Sm, Lu, Yb$)的表面水合作用,深入研究了表面离子配位行为和优先吸附位点^[8]。

尽管相关结构解析技术实现了一种晶体材料或一类材料的局域结构解析,然而,快速、批量地获取结构信息仍充满挑战。另外,晶体中引入稀土离

子时,由于稀土离子与被取代离子的物理化学性质差异,与原晶格结构相比,稀土离子的局域配位环境,如对称性、构型、键长等信息通常会发生改变。目前,相关局域结构信息的完整、全面解析仍存在一定的难度。考虑到在等价取代的情况下,阴离子配体、对称性和多面体形状等通常保持不变,由于离子半径等差异,仅稀土离子与阴离子配体之间的键长发生变化。已有研究表明,键长变化通过晶体场分裂与质心位移两种物理机制,共同决定了发光离子的能级位置,进而调控其发光波长与效率。键长缩短会显著增强晶体场分裂,而键长增加则通过提升阴离子极化率增大质心位移。在发光材料设计中,可通过精确调控键长与配位环境,实现对发光性能的定向优化^[9-10]。因此,研究稀土离子局域结构键长这一参量不仅重要,而且极大地降低了研究的难度。针对键长的研究,Dorenbos等提出使用离子半径差值的0.6倍作为键长变化量来预测由掺杂引起的稀土离子第一配位壳层键长及其变化^[11-13],该方法虽然可以快速对键长进行评估,但得到的结果过于粗糙。与之相比,目前的前沿实验表征和模拟技术可以给出更精确的结果^[4,14],但评估过程异常耗时。

近年来,机器学习尤其是深度学习因其高效、准确的预测能力引起了人们极大的兴趣^[15-16]。Hansen等开发了“键袋”(Bag of bond)方法来预测分子的原子化能和电子性质,发现即使采用简单对偶力场(Pairwise force fields),其预测精度也可以与使用杂化泛函的模拟计算的精度相媲美,且耗时更少^[17]。基于模拟计算的数据集,Pilania等使用核岭回归准确预测了多种材料体系的性质^[18];而Zhuo等利用支持向量回归模型准确预测了材料的德拜温度^[19]。模拟计算和机器学习算法的结合为发现和筛选高性能无机荧光粉提供了新的研究方法。然而,利用机器学习预测稀土离子局域结构键长却鲜有报道。

本文采用基于密度泛函的第一性原理计算,研究了稀土离子进入晶格后其第一配位壳层键长变化的影响因素,主要包括稀土离子半径、配位数、阴离子配体以及多面体形状等。基于第一性

原理计算数据集,采用反向传输神经网络算法和支持向量机等模型进行训练,并成功预测晶体材料中稀土离子的局域配位键长,相关研究结果为稀土掺杂无机晶体材料的局域结构信息的快速、准确预测提供了参考。

2 实验方法

本文利用基于密度泛函(DFT)的第一性原理模拟计算包VASP^[20-23]对稀土掺杂的晶体进行模拟计算,掺杂的离子包括 $\text{La}^{3+} \sim \text{Lu}^{3+}$ 、 Y^{3+} 等三价稀土离子(RE^{3+}),而晶体基质以氧化物、氟化物、硫化物和氮化物等无机材料为主,同时满足稀土离子等价取代。等价取代可以有效避免因复杂的电荷补偿机制而产生的复杂结构模型,以简化计算。本研究采用的晶体都是较为成熟的晶体,且其结构均是已公开发表的,相关晶体信息均可在ICSD数据库获取。以获取的晶体结构为初始模型,进行弛豫计算,采用投影缀加平面波(PAW)赝势和Perdew-Burke-Ernzerhof(PBE)交换泛函描述粒子间的相互作用^[24-25]。平面波展开的截断能设为520 eV,并使用 $1 \times 1 \times 1$ Gamma k 点网格对第一布里渊区进行抽样,以确保弛豫精度达到 10^{-5} eV。优化超胞内的原子内坐标直至单个原子上的力小于 10 eV/nm^3 ,所有计算均采用自旋极化。构建的晶体超胞足够大,且掺杂元素位于超胞中心。若掺杂离子不在超胞中心,则需要重构晶胞及超胞。此外,若超胞中原子数超过300,受限于服务器算力,我们会构建根号倍超胞而非整数倍超胞,保证计算速度的同时,使得掺杂离子间的相互作用尽可能的小。另外,为了加速计算,除La和Y外(其价电子分别为 $5s^2 5p^6 5d^1 6s^2$ 和 $4s^2 4p^6 4d^2 5s^1$),其他元素均采用了冻结部分f电子的RE_3赝势。需要指出的是,本研究中的计算都在相同或相近的条件下进行,使得训练的数据能够统一、对齐,避免了额外的干扰。

我们采用反向传输神经网络算法(BPNN)^[26-27],其输入层节点数为22,有1个隐藏层,其节点数根据经验公式 $\text{hiddennum} = \sqrt{m+n} + a$ 确定,其中 m 为输入层节点数, n 为输出层节点数, a 是任意值,通常取1~10。因此,隐藏层节点数的取值范围为6~15。输出层节点数为1,激活函数采用Log-sigmoid函数,学习速率为0.01,训练目标最小均方误差设置为0.000 1,最大迭代次数设置为1 000,构建并训练了预测模型,预测无机晶体材

料中稀土离子第一配位壳层的键长以及键长随稀土离子半径变化的斜率。此外,我们还使用支持向量回归(SVR)模型进行训练和对比,SVR的核函数类型为径向基函数(RBF),采用网格搜索法进行惩罚参数 C 和核系数 γ 的优化,以指数形式在 $-10 \sim 10$ 、每间隔0.5遍历 C 和 γ 的组合,选出最优超参数以提升模型性能。以16种稀土离子掺杂44种基质晶体,即704种稀土掺杂无机晶体的键长作为数据集,并选取22个特征参数,主要包括配位数、稀土电负性(Pauling标度)、配体电负性(Pauling标度)等(见补充文件表S1~S2)。数据集按适当比例划分为训练集和测试集,通过算法揭示键长或斜率与特征参量之间的关联关系,使用平均绝对误差(MAE)、均方误差(MSE)、均方根误差(RMSE)和决定系数(R^2)来评估训练的效果。随后,利用Shapley值分析各特征参量的贡献^[28]。基于训练好的模型,成功预测了96种稀土掺杂晶体的键长,并通过计算验证了其可靠性。

3 结果与讨论

本文首先研究了稀土离子半径对局域配位键长的影响。图1和表S3展示了 RE^{3+} 掺杂的九种代表性晶体(LaCl_3 、 La_2S_3 、 LaSi_3N_5 、 LaBO_3 、 LaPO_4 、 La_2O_3 、 $\text{Y}_3\text{Al}_5\text{O}_{12}$ (YAG)、 LiYF_4 (YLF)和 LuBO_3)的第一配位壳层的平均键长,结果显示,稀土离子第一配位壳层的平均键长随稀土离子半径的增大而增加,两者呈线性相关。由于包含了氧化物、卤化物、硫化物和氮化物晶体,这种线性关系在一定程度上被认为是普适的。结果表明,平均键长线性依赖于稀土离子半径。离子半径越大,平均键长越长,

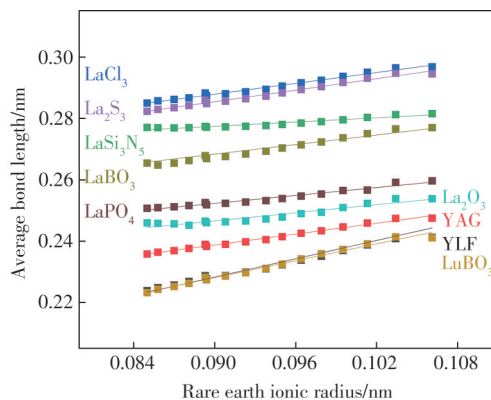


图1 第一配位层中 RE^{3+} 的平均键长随稀土离子半径的变化

Fig.1 Average bond length of RE^{3+} of the first coordination shell versus the rare earth ionic radius

这与实验结果一致^[4]。需要指出的是,从模拟计算结果中可以提取出稀土离子第一配位壳层的所有键长。然而,由于配位数和多面体形状的差异,对每个键长逐一进行比较是非常困难的。因此,本文优先考虑并采用平均键长进行探讨。

此外,如图 1 所示,不同晶体表现出不同的斜率和截距。例如, $RE^{3+}:\text{LuBO}_3$ 的斜率和截距分别为 0.919 和 1.456。对于 YAG 和 LaPO_4 , 它们的斜率和截距分别为 0.592 和 1.857, 0.458 和 2.111。随着配位数(CN)从 6CN 增加到 8CN、9CN, 斜率值逐渐减小而截距逐渐增大。同时,斜率和截距还取决于阴离子配体。可以看出,在 LaCl_3 、 La_2S_3 、YAG 和 YLF 晶体中,随着配体从 Cl^- 、 S^{2-} 、 O^{2-} 变化到 F^- , 其斜率和截距值也发生了变化。此外, La_2O_3 、 LaPO_4 和 LaBO_3 中稀土离子的配位多面体形状不同,导致了不同的斜率和截距。因此,除稀土离子半径外,配位数、配体和多面体形状也会对键长产生影响^[29-31]。我们将在下文逐一研究配位数、配体和多面体形状对稀土离子平均键长的影响。

我们选择具有相同配体但配位数不同的氧化物晶体来研究配位数对稀土离子第一配位壳层的平均键长的影响。所选晶体包括 $\text{La}_4\text{Al}_2\text{O}_9$ (5CN)、 ScPO_6 (5CN)、 LuBO_3 (6CN)、 $\text{La}_4\text{Al}_2\text{O}_9$ (6CN)、 $\text{La}_2\text{Si}_2\text{O}_7$ (6CN)、 La_3PO_7 (7CN)、 Y_2SiO_5 (7CN)、 $\text{Lu}_2\text{Si}_2\text{O}_7$ (7CN)、 Lu_2SiO_5 (7CN)、YAG (8CN)、 YBO_3 (8CN)、 LuPO_4 (8CN)、 LaP_3O_9 (8CN)、 YB_5O_9 (9CN)、 LaBO_3 (9CN)、 LaPO_4 (9CN)、 GdB_3O_6 (10CN)、 LaB_3O_6 (10CN) 和 $\text{La}_2\text{B}_8\text{O}_{15}$ (11CN) 等, 配位数为 5~11。图 2 和表 S1 展示了部分具有代表性的稀土掺杂无机晶体的平均键长。结果显示,配位数不同,键长的斜率和截距不同,即配位数会对键长产生影响。图 3(a) 展示了不同配位数下晶

体的斜率和截距,可以看出,随配位数增加,键长的斜率线性减小,而截距呈线性增加趋势。基于斜率和截距,结合稀土离子半径,即可评估和预测稀土掺杂无机晶体的平均键长。然而,在相同配位数下,斜率差在 0.03~0.12 之间波动,截距差在 0.001~0.20 之间波动。如此大的波动会导致键长预测值与准确值之间存在较大的偏差,甚至造成预测失效。这表明,除了配位数,还存在其他影响因素。

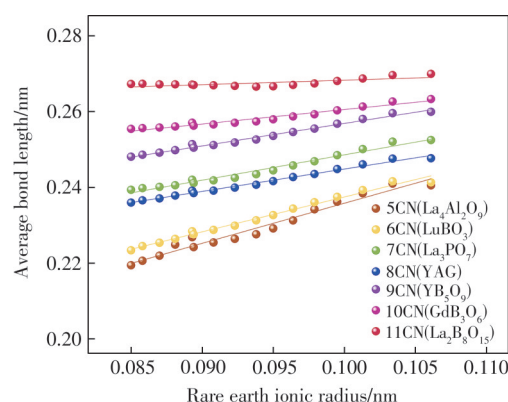


图 2 不同配位数下 RE^{3+} 的平均键长随稀土离子半径的变化

Fig.2 Average bond length of RE^{3+} in different coordination numbers versus the rare earth ionic radius

在研究其他影响因素之前,需要强调的是,随着配位数的增加,多面体逐渐接近于球形。从图 3(b) 中可以看出,与 $\text{La}_4\text{Al}_2\text{O}_9$ 晶体中的五配位畸变四方锥相比, LuBO_3 晶体中的八面体明显多出一个顶点。随着顶点或配位数逐渐增加,多面体逐渐从四方锥演变为二十面体,同时体积增大。体积增大后的局域配位环境更适合尺寸较大的稀土离子占据,间接导致键长斜率随着配位数的增加而减小。从这个角度看,可以假设具

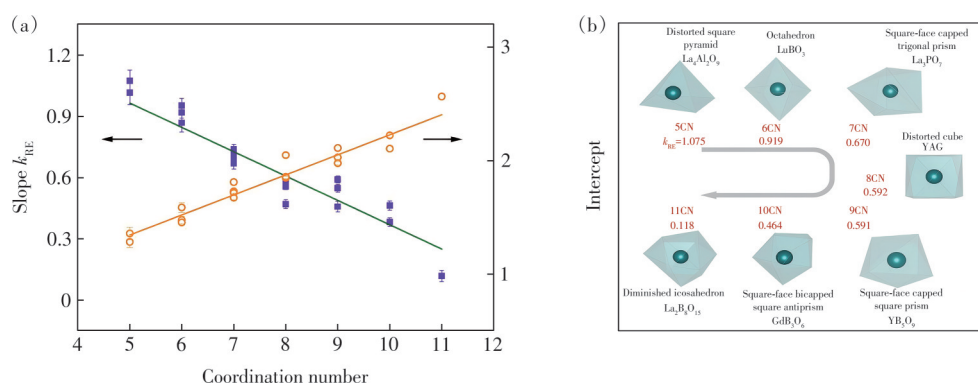


图 3 (a) 依赖于配位数的斜率 k_{RE} 和截距; (b) 配位数 5~11 的稀土离子周围的多面体形状

Fig.3 (a) The coordination number dependent slopes k_{RE} and intercepts. (b) Polyhedral shapes around the rare earth ions with coordination number ranging from 5 to 11

有相同配位数的晶体其多面体通常相似,进而在一定程度上具有相似的斜率。配位数的改变也反映了配位多面体以及晶体结构的演变,特别是在结晶过程中原子和分子的堆叠,需要进一步研究。

接下来,我们选择了两组晶体:YN、YP、YAs和YSb,以及 K_3YF_6 、 K_3YCl_6 、 K_3YBr_6 、 K_3YI_6 ,研究阴离子配体对键长的影响。YN、YP、YAs和YSb晶体(空间群 $Fm-3m$)中阳离子的配位数均为6CN,多面体的形状也相同,为八面体配位结构,而 K_3YF_6 、 K_3YCl_6 、 K_3YBr_6 和 K_3YI_6 的晶体结构、阳离子的配位数和配位多面体形状均相同。因此,选择的两组晶体,阴离子配体是唯一的变量。图4展示了平均键长随稀土离子半径的变化规律。可以看出,随着配体从N、P、As变化到Sb,稀土离子的平均键长增加,截距随之提高的同时,斜率从0.65降至

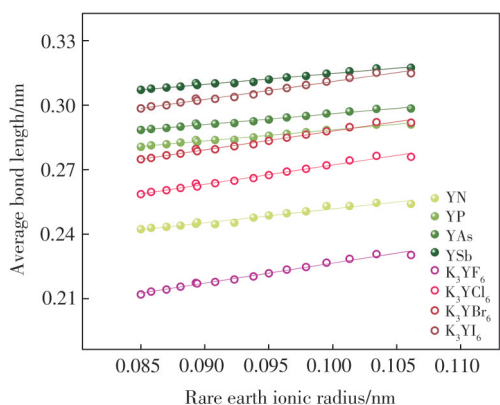
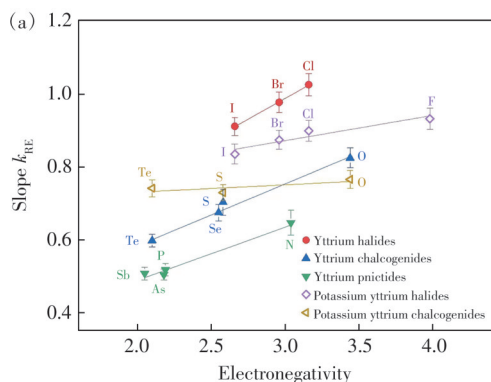


图4 不同配体下 RE^{3+} 的平均键长随稀土离子半径的变化

Fig.4 Average bond length of RE^{3+} with different ligands versus the rare earth ionic radius



0.50。在 K_3YF_6 、 K_3YCl_6 、 K_3YBr_6 、 K_3YI_6 晶体中观察到类似的变化趋势,结果显示,同族阴离子配体对稀土离子的平均键长产生显著影响^[30]。

除了同族阴离子,同周期阴离子配体的晶体,稀土离子的平均键长、斜率和截距也不同。例如,YN与 K_3YF_6 、YP与 K_3YCl_6 ,平均键长的斜率和截距都不相同。图5展示了同族和同周期阴离子电负性与斜率的关系,清晰地展示了配体对稀土离子平均键长、斜率的影响。除前述两组晶体外,还包含了钇卤化物、钇硫化物、钾钇硫化物作为对比。由于 YF_3 中不存在6CN多面体,且 $KYSe_2$ 不稳定,故未包含这两种晶体。图5展示了稀土离子平均键长的斜率随阴离子电负性的变化规律。图5(a)具体展示了具有同族阴离子配体的晶体其斜率的变化。可以看出,由同族阴离子形成的晶体其斜率值随着阴离子电负性的增加而线性增加,卤族晶体的斜率最大,含氮族元素晶体的斜率最小。卤族元素的电负性大于相应的氧族和氮族元素^[32],其产生的共价效应可能会对键长产生影响^[33]。此外,与二元化合物相比,具有同族阴离子配体的三元化合物其斜率变化更缓慢,表明晶体中包含其他的元素可以减缓平均键长的变化。图5(b)展示了具有同周期阴离子配体的晶体其电负性与斜率的关系。可以看出,随电负性增加,斜率值线性增加。配体的电负性越小,其形成共价键而非离子键的可能性越大。共价键特性在一定程度上阻碍了离子的移动,造成相对较小的键长变化。综上,阴离子配体显著影响稀土离子的平均键长。

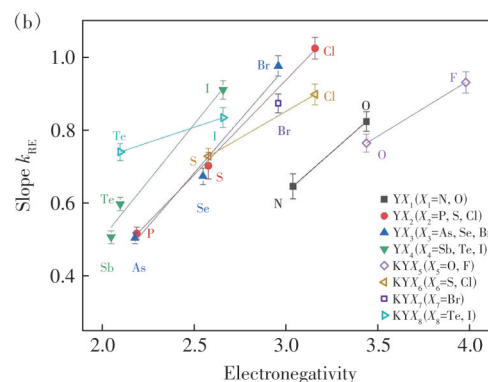


图5 斜率 k_{RE} 与同族(a)及同周期(b)阴离子电负性的关系。YO代表 Y_2O_3 ; YCl、YBr、YI分别代表 YCl_3 、 YBr_3 、 YI_3 ; KYO、KYS、KYTe分别代表 KYO_2 、 KYS_2 、 $KYTe_2$; KYF、KYCl、KYBr、KYI分别代表 K_3YF_6 、 K_3YCl_6 、 K_3YBr_6 、 K_3YI_6

Fig.5 The slope k_{RE} versus electronegativity of anions in the same group (a) and the same period (b). The YO means Y_2O_3 ; YCl, YBr and YI the YCl_3 , YBr_3 and YI_3 ; KYO, KYS and KYTe the KYO_2 , KYS_2 and $KYTe_2$; KYF, KYCl, KYBr and KYI the K_3YF_6 , K_3YCl_6 , K_3YBr_6 and K_3YI_6 , respectively

最后,我们还研究了配位多面体形状对稀土离子平均键长的影响。如图 6 所示,配位数从 6CN 到 10CN,多面体形状相近时,稀土离子平均键长的斜率差异较小。例如, $\text{La}_4\text{Al}_2\text{O}_9$ 和 LuBO_3 中的六配位八面体的斜率值分别为 0.953 和 0.919,两者相差 0.034。 LuPO_4 和 LaP_3O_9 中的八配位三角十二面体的斜率值分别为 0.557 和 0.470,两者

的差异约为 0.09。配位数(八配位)相同、多面体形状不同时,斜率的差异略有增加,但不超过 0.15。与前述影响因素相比,配位多面体形状对键长的影响较小。基于上述讨论可知,配位数、稀土离子半径和阴离子配体对进入晶格后的稀土离子第一配位壳层的平均键长影响较大,与之相比,多面体形状对键长的影响较小。

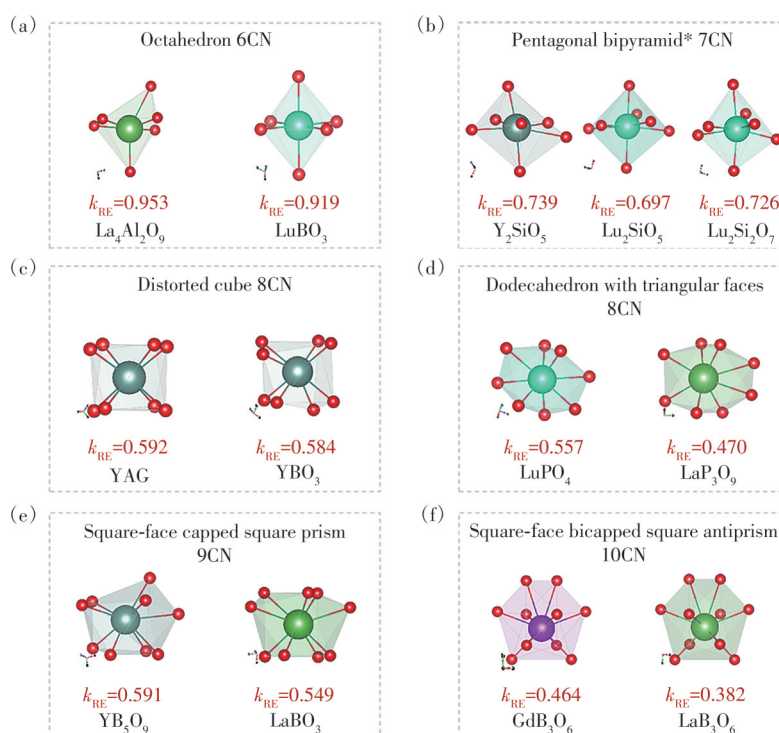


图 6 多面体形状与斜率 k_{RE} 的比较。(a)~(f)八面体、五角双锥、立方体、三角面十二面体、方帽四方柱和双方帽四方反棱柱,符号*表示 Y_2SiO_5 和 Lu_2SiO_5 中的五角双锥是畸变的

Fig.6 Comparison of the polyhedral shapes and the slopes of k_{RE} . (a)~(f) Octahedron, pentagonal bipyramid, cube, dodecahedron with triangular faces, square-face capped square prism and square-face bicapped square antiprism, respectively. The symbol * means that the pentagonal bipyramid is distorted in Y_2SiO_5 and Lu_2SiO_5

需要指出的是,上述研究考虑的影响因素非常有限,且量化每个影响因素的具体贡献也较为困难。由于机器学习可以同时考虑多个影响因素,并能直接预测,因此,本文还采用机器学习研究了稀土离子第一配位壳层平均键长。首先,我们学习并预测了稀土离子平均键长的斜率 k_{RE} ,并与计算结果进行比较。以 44 种稀土离子掺杂的基质晶体的 44 个 k_{RE} 值为数据集,以阴离子配体、配位数以及与晶体结构和化学元素相关的密度和晶胞参数等 16 个参数为特征参量。数据集按 6:4 的比例划分为训练集和测试集,采用 BPNN 算法进行学习训练,并与 SVR 结果进行比较。如图 S1 所示,两种算法的决定系数均为 ~0.8,表明斜率 k_{RE} 的预测存在一定的偏差,造成偏差的原因可能

是样本量小且特征参量数量较多,这两者导致训练和预测的不确定性增加。此外,我们利用 Shapley 值分析了各特征参数对稀土离子平均键长斜率 k_{RE} 的贡献^[28],结果显示,配体电负性、晶体密度和晶格参数对斜率的影响最大,而配位数的影响相对较小,与第一性原理计算的结果不同。通常而言,阴离子配体和配位数对 k_{RE} 的影响较大,而晶体密度和晶格参数的影响较小。机器学习的研究结果表明,该方法能够挖掘出更多的影响因素。

在此基础上,本文以 16 种稀土离子掺杂到 44 种基质晶体中的 704 条键长为数据集,以配位数、阴离子配体、稀土离子特性等 22 个参数为特征参量,使用 BPNN 和 SVR 进行训练和预测。需要指出的是,已有研究结果显示,在特征参量数目

远低于样本数的情况下,特征相关性分析引起的参量数目微调对模型训练和评估能力的影响较小或可以忽略^[34-37],本文直接进行了学习和训练。稀土掺杂无机晶体的结构参数列于表 1 和表 S2,相关的键长数据列于表 S3。数据集按 8:2 的比例随机划分为训练集和测试集。图 7(a)为 BPNN 算法的预测值,较小的 MAE(0.023 1)、MSE(0.000 9)和 RMSE(0.030 1)值,以及高达 0.994 0 的决定系数 R^2 ,表明 BPNN 算法预测的稀土离子键长与模拟计算的结果高度吻合。为了避免过拟合,我们还随机选择一种稀土离子掺杂的晶体的数据作为验证集,其余作为训练集,应用十六折交叉验证方

法(见补充文件),训练结果如图 S2~S3 所示。结果显示,训练预测的结果与前述结果吻合,因此,可以排除过拟合。作为对比,我们还采用 SVR 算法进行了训练。需要指出的是,在训练过程中使用随机划分的十六折交叉验证,以便与 BPNN 训练结果及其验证评估方法比较和对照。图 7(b)展示了 SVR 的训练结果,MAE、MSE、RMSE 和 R^2 值与 BPNN 的训练和评估结果均相近,表明无论是 SVR 还是 BPNN 其模型的预测和泛化能力均较好。由于训练模型采用的 22 个特征参量均可从已有手册中获取,根据特征参量和训练的模型可以直接预测稀土离子的局域配位结构键长。

表 1 稀土掺杂晶体的结构信息,包括空间群、配位数、阴离子配体、多面体形状和 k_{RE}

Tab. 1 The structure information, including space group, coordination number, anionic ligand, polyhedral shape and k_{RE} of the rare earth-doped crystals

Rare earth-doped crystal	Space group	Coordination number	Anionic ligand	Polyhedral shape	k_{RE}
La ₂ O ₃	<i>P-3m1</i>	7	O	Face-capped octahedron	0.439
La ₂ S ₃	<i>Pnma</i>	7	S	Square-face capped trigonal prism	0.627
YLF	<i>I41/a</i>	8	F	Dodecahedron with triangular faces	1.006
LaSi ₃ N ₅	<i>P2₁-2₁-2₁</i>	9	N	Tricapped triangular prism	0.232
La ₄ Al ₂ O ₉	<i>P1</i>	5	O	Square pyramid	1.075
ScPO ₆	<i>P2₁/c</i>	5	O	Triangular bipyramidal	1.016
La ₄ Al ₂ O ₉	<i>P1</i>	6	O	Octahedron	0.953
LuBO ₃	<i>R3ch</i>	6	O	Octahedron	0.919
La ₂ Si ₂ O ₇	<i>P2₁/c</i>	6	O	Pentagonal pyramid	0.869
Y ₂ SiO ₅	<i>C2/c</i>	7	O	Pentagonal bipyramid	0.739
Lu ₂ Si ₂ O ₇	<i>P4₁2₁2</i>	7	O	Pentagonal bipyramid	0.726
Lu ₂ SiO ₅	<i>P2₁/c</i>	7	O	Pentagonal bipyramid	0.697
La ₃ PO ₇	<i>Cm</i>	7	O	Square-face capped trigonal prism	0.670
YAG	<i>Ia-3</i>	8	O	Cube	0.592
YBO ₃	<i>P11m</i>	8	O	Cube	0.584
LaF ₃	<i>P-3c1</i>	9	F	Tricapped triangular prism	0.608
LaCl ₃	<i>P6₃/m</i>	9	Cl	Tricapped triangular prism	0.615
LaBr ₃	<i>P6₃/m</i>	9	Br	Tricapped triangular prism	0.606
Y ₂ O ₃	<i>Fm-3m</i>	6	O	Octahedron	0.824
YS	<i>Fm-3m</i>	6	S	Octahedron	0.702
YSe	<i>Fm-3m</i>	6	Se	Octahedron	0.674
YTe	<i>Fm-3m</i>	6	Te	Octahedron	0.597
Na ₃ Lu(MoO ₄) ₄	<i>I4₁/a</i>	8	O	Tetragonal	—
YAl ₃ (BO ₃) ₄	<i>R3₂</i>	6	O	Trigonal	—
LuVO ₄	<i>I4₁/amd</i>	8	O	Tetragonal	—
LuPO ₄	<i>I4₁/amd</i>	8	O	Dodecahedron with triangular	0.557
LaP ₃ O ₉	<i>C222₁</i>	8	O	Dodecahedron with triangular	0.470
YB ₅ O ₉	<i>P2₁/c</i>	9	O	Square-face capped	0.591
LaBO ₃	<i>Pnma</i>	9	O	Square-face capped square prism	0.647
LaPO ₄	<i>P2₁/c</i>	9	O	Tricapped triangular prism	0.458
GdB ₃ O ₆	<i>C2/c</i>	10	O	Square-face bicapped square antiprism	0.464
LaB ₃ O ₆	<i>C2/c</i>	10	O	Square-face bicapped square antiprism	0.382
La ₂ B ₈ O ₁₅	<i>P2/c</i>	11	O	Diminished icosahedron	0.118
YCl ₃	<i>C2/m</i>	6	Cl	Octahedron	1.025

表 1 (续)

Rare earth-doped crystal	Space group	Coordination number	Anionic ligand	Polyhedral shape	k_{RE}
YBr ₃	<i>C2/m</i>	6	Br	Octahedron	0.977
YI ₃	<i>R-3</i>	6	I	Octahedron	0.91
YN	<i>Fm-3m</i>	6	N	Octahedron	0.646
YP	<i>Fm-3m</i>	6	P	Octahedron	0.517
YAs	<i>Fm-3m</i>	6	As	Octahedron	0.503
YSb	<i>Fm-3m</i>	6	Sb	Octahedron	0.506
K ₃ YF ₆	<i>Fm-3m</i>	6	F	Octahedron	0.931
K ₃ YCl ₆	<i>Fm-3m</i>	6	Cl	Octahedron	0.898
K ₃ YBr ₆	<i>Fm-3m</i>	6	Br	Octahedron	0.874
K ₃ YI ₆	<i>Fm-3m</i>	6	I	Octahedron	0.834
KYO ₂	<i>R-3m</i>	6	O	Octahedron	0.764
KYS ₂	<i>R-3m</i>	6	S	Octahedron	0.728
KYTe ₂	<i>R-3m</i>	6	Te	Octahedron	0.740
YPO ₄	<i>I4₁/amd</i>	8	O	Tetragonal	—
NaGdGeO ₄	<i>Pnma</i>	6	O	Orthorhombic	—
La ₂ (WO ₄) ₃	<i>C2/c</i>	8	O	Monoclinic	—

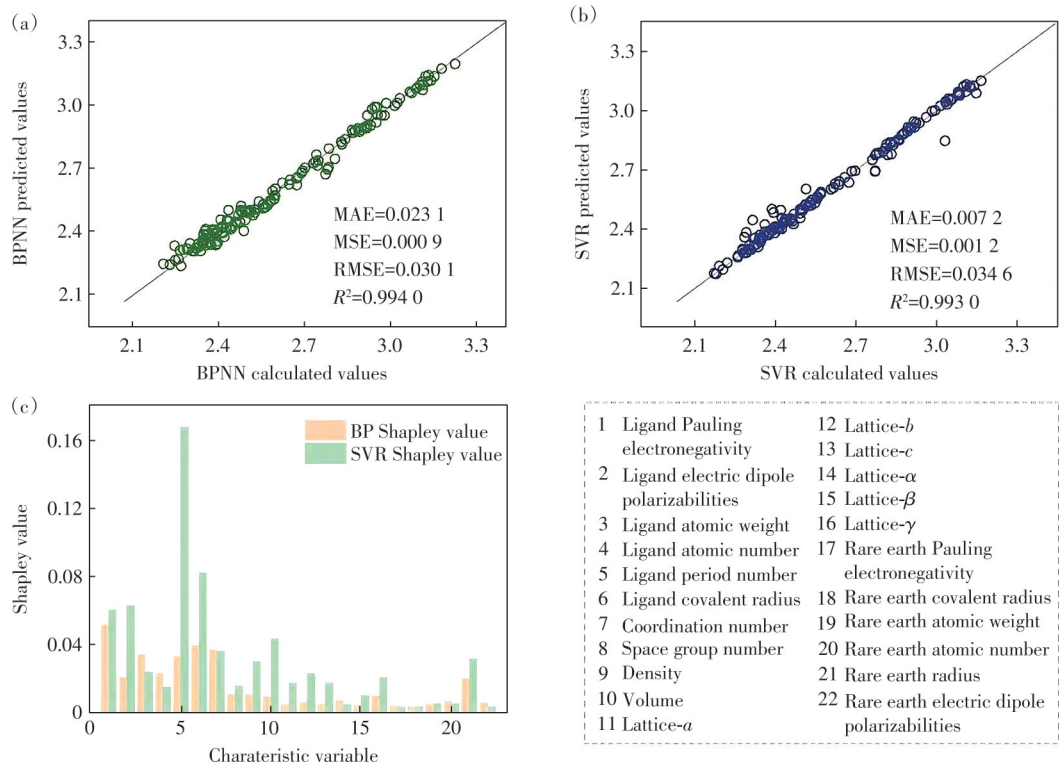


图 7 稀土掺杂晶体的学习模型。BPNN(a)和SVR(b)结果;(c)BPNN 和SVR 中各特征的Shapley 值比较,同时列出了特征参数

Fig.7 The learned models in rare earth-doped crystal. BPNN(a) and SVR(b) results. (c)Comparison of Shapley values for each feature in BPNN and SVR. The characteristic features were also listed

考虑到基于机器学习的预测方法大多依赖复杂的模型,这些模型的核心特点是通过大量参数拟合高维特征间的非线性关系,其内部的决策逻辑往往涉及特征间的交互作用、非线性转换等复杂模式,通常无法简化为明确的解析表达式。为了弥补这一不足,我们采用Shapley 值量化每

个输入特征参量的具体贡献^[28]。这种分析方式虽然不能替代解析表达式,但能直观展示特征与预测结果之间的关联逻辑,便于学习和解释模型行为,也为进一步特征的优化、预测逻辑的改进提供了可量化的参考。从图7(c)中可以看出,各特征参量的贡献,尤其是Shapley 值较大的参量,其

贡献并不总是一致。BPNN算法的结果显示,配体周期数、配体共价半径、配体Pauling电负性、配体原子电偶极极化率、配位数和稀土离子半径等的贡献较大;与之相比,SVR的结果显示,除了配体Pauling电负性、配体共价半径等与配体有关的参量以及配位数、稀土离子半径等,晶体密度、晶胞体积和晶胞参数等也有相当的贡献,与BPNN的结果不同。尽管如此,两种算法呈现出的配体性质、配位数以及稀土离子半径的贡献均较为显著,与第一性原理结果一致。如前所述,由于第一性原理方法的局限性,我们仅研究了稀土离子半径、配位数、阴离子配体和配位多面体形状等影响因素。然而,机器学习可以同步研究多个特征参量的影响,并量化其贡献。更重要的是,机器学习的训练和预测速率快、耗时短,展现出一定的效率优势。总而言之,我们通过机器学习和Shapley值分析,研究了晶格膨胀或收缩的影响因素。稀土离子进入晶格后会与基质晶体相互作用,由于稀土离子与基质离子之间的物理化学性

质差异,晶体的局域配位环境会不可避免地发生变化。阴离子配体、配位数、晶胞体积、晶格参数和稀土离子半径共同决定了局域晶格畸变的程度及其结构的稳定性,并通过机器学习算法实现稀土离子第一配位壳层键长的快速、高效、精确预测。

为了进一步验证训练模型的可靠性,我们以22个特征参量为输入(表S2),预测了16种稀土离子掺杂 $\text{Na}_5\text{Lu}(\text{MoO}_4)_4$ 、 $\text{YAl}_3(\text{BO}_3)_4$ 、 LuVO_4 、 YPO_4 、 NaGdGeO_4 和 $\text{La}_2(\text{WO}_4)_3$ 六个未知晶体的平均键长。图8(a)、(b)展示了预测结果与计算结果的比较,其中计算结果为采用第一性原理计算的平均键长(表S3)。从评估指标可以看出,BPNN算法预测和计算结果吻合较好,与之相比,SVR模型的决定系数较小,仅为0.8182,表现出相对较差的预测性能。这可能与BPNN算法具有更强的泛化能力有关,同时,减小特征参量的数量有望进一步提升训练模型的可靠性和准确性,相关研究正在进行中。

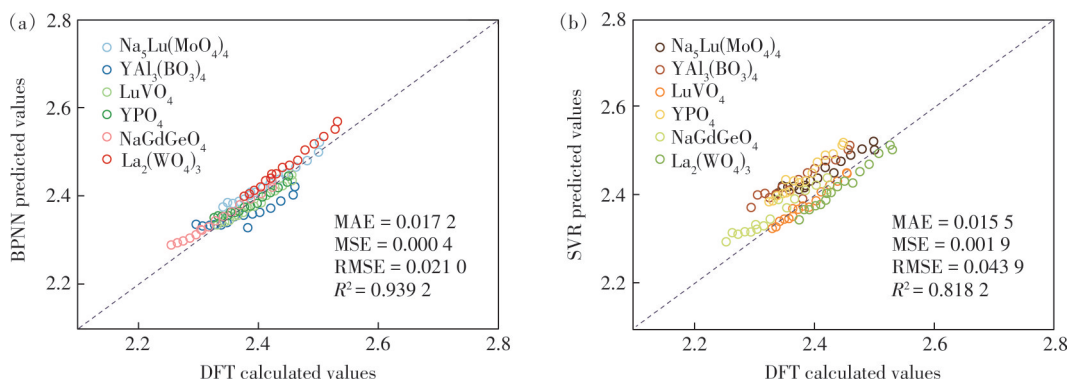


图8 96种未报道稀土掺杂晶体的预测键长与计算键长的比较。(a)使用学习好的BPNN模型预测;(b)使用学习好的SVR模型预测

Fig.8 Comparison of the predicted and calculated bond lengths of 96 rare earth-doped unreported crystals, the prediction with the learned BPNN(a) and SVR(b) models

4 结 论

本文采用第一性原理计算研究了稀土离子第一配位壳层的平均键长,发现键长与稀土离子半径呈线性依赖关系;同时,配位数和阴离子配体对平均键长有显著影响,而配位多面体形状的影响相对较小。结合机器学习BPNN和SVR算法,研究了22个特征参量对平均键长的影响。训练结果显示,与阴离子配体相关的特征参量对稀土离子平均键长的贡献较大,采用Shapley值还可以量化每一个特征参量的贡献,并实现对未

知稀土掺杂无机晶体局域结构的可靠预测。本研究展示了一种简单、快速预测稀土掺杂无机晶体材料局域结构键长的方法,尽管存在非数据驱动的深度学习预测,但该方法为完整、全面的局域结构信息的批量预测提供了一个可能的途径,也为按需设计和优化稀土掺杂无机晶体材料奠定了基础。

本文补充文件和专家审稿意见及作者回复内容的下载地址:<http://cjl.lightpublishing.cn/thesisDetails#10.37188/CJL.20250215>

参 考 文 献:

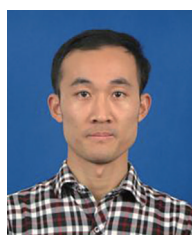
- [1] WANG D G, LIU X F, LAI Y T, *et al.* High-power and high-energy coherent sources at 1.1- μm based on compact diode-end-pumped Nd:YAG crystal fiber lasers [J]. *Infrared Phys. Technol.*, 2024, 136: 105113.
- [2] MIAO J G, PAN Y Z, QU S L. Comparison of small-scale actively and passively Q-switched eye-safe intracavity optical parametric oscillators at 1.57 μm [J]. *Chin. Phys. Lett.*, 2012, 29(7): 074201.
- [3] 徐军, 苏良碧, 徐晓东, 等. 激光晶体的现状及发展趋势 [J]. *无机材料学报*, 2006, 21(5): 1025-1030.
XU J, SU L B, XU X D, *et al.* Recent developments and research frontier of laser crystals [J]. *J. Inorg. Mater.*, 2006, 21(5): 1025-1030. (in Chinese)
- [4] POPOV V V, MENUSHENKOV A P, YASTREBTSSEV A A, *et al.* Multiscale study on the formation and evolution of the crystal and local structures in lanthanide tungstates $\text{Ln}_2(\text{WO}_4)_3$ [J]. *J. Alloys Compd.*, 2022, 910: 164922.
- [5] POPOV V V, MENUSHENKOV A P, YASTREBTSSEV A A, *et al.* Effects of synthesis conditions on the crystal and local structures of high-entropy oxides $\text{Ln}_2\text{M}_2\text{O}_7$ ($\text{Ln}=\text{La}-\text{Yb}$, Y ; $\text{M}=\text{Ti}$, Zr , Ce) [J]. *Ceram. Int.*, 2024, 50(3): 5319-5335.
- [6] KRIEKE G, DOKE G, ANTUZEVICI A, *et al.* Structure and persistent luminescence of novel Pr-doped $\text{Mg}_3\text{Lu}_2\text{Ge}_3\text{O}_{12}$ garnet [J]. *J. Alloys Compd.*, 2023, 957: 170421.
- [7] ZHAI Y L, YANG X W, ZHAO S N, *et al.* Directly imaging of the atomic structure of luminescent centers in $\text{CaYAlO}_4:\text{Ce}^{3+}$ [J]. *J. Nano Res.*, 2023, 16(2): 3004-3009.
- [8] KHALILI R, LARSSON A C, TELKKI V V, *et al.* Local structures of rare earth phosphate minerals by NMR [J]. *J. Solid State Chem.*, 2022, 311: 123097.
- [9] DORENBOS P. 5d-level energies of Ce^{3+} and the crystalline environment. I. Fluoride compounds [J]. *Phys. Rev. B*, 2000, 62(23): 15640-15649.
- [10] DORENBOS P. 5d-level energies of Ce^{3+} and the crystalline environment. II. Chloride, bromide, and iodide compounds [J]. *Phys. Rev. B*, 2000, 62(23): 15650-15659.
- [11] DORENBOS P. Relation between Eu^{2+} and Ce^{3+} f $\leftrightarrow$ d-transition energies in inorganic compounds [J]. *J. Phys.: Condens. Matter*, 2003, 15(27): 4797-4807.
- [12] ZHUO Y, HARIYANI S, YOU S H, *et al.* Machine learning 5d-level centroid shift of Ce^{3+} inorganic phosphors [J]. *J. Appl. Phys.*, 2020, 128(1): 013104.
- [13] DORENBOS P. A review on how lanthanide impurity levels change with chemistry and structure of inorganic compounds [J]. *ECS J. Solid State Sci. Technol.*, 2013, 2(2): R3001-R3011.
- [14] SAIDI F, MOKHDAR S, DERGAL M, *et al.* *Ab initio* study of the structural, electronic, magnetic, mechanical, optical, and dynamical properties of the rare-earth dihydrides MH_2 ($\text{M}=\text{Yb}$, Sc , Eu , Y , Lu and Gd) [J]. *Vacuum*, 2023, 212: 112011.
- [15] MAHESH B. Machine learning algorithms—a review [J]. *Int. J. Sci. Res.*, 2020, 9(1): 381-386.
- [16] BZDOK D, ALTMAN N, KRZYWINSKI M. Statistics versus machine learning [J]. *Nat. Methods*, 2018, 15(4): 233-234.
- [17] HANSEN K, BIEGLER F, RAMAKRISHNAN R, *et al.* Machine learning predictions of molecular properties: accurate many-body potentials and nonlocality in chemical space [J]. *J. Phys. Chem. Lett.*, 2015, 6(12): 2326-2331.
- [18] PILANIA G, WANG C C, JIANG X, *et al.* Accelerating materials property predictions using machine learning [J]. *Sci. Rep.*, 2013, 3: 2810.
- [19] ZHUO Y, TEHRANI A M, OLIYNYK A O, *et al.* Identifying an efficient, thermally robust inorganic phosphor host via machine learning [J]. *Nat. Commun.*, 2018, 9(1): 4377.
- [20] KRESSE G, HAFNER J. *Ab initio* molecular dynamics for liquid metals [J]. *Phys. Rev. B*, 1993, 47(1): 558-561.
- [21] KRESSE G, HAFNER J. *Ab initio* molecular-dynamics simulation of the liquid-metal-amorphous-semiconductor transition in germanium [J]. *Phys. Rev. B*, 1994, 49(20): 14251-14269.
- [22] KRESSE G, FURTHMÜLLER J. Efficiency of *ab-initio* total energy calculations for metals and semiconductors using a plane-wave basis set [J]. *Comput. Mater. Sci.*, 1996, 6(1): 15-50.
- [23] KRESSE G, FURTHMÜLLER J. Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set [J]. *Phys. Rev. B*, 1996, 54(16): 11169-11186.

- [24] HAN T P J, JONES G D, SYME R W G. Site-selective spectroscopy of Nd^{3+} centers in $\text{CaF}_2:\text{Nd}^{3+}$ and $\text{SrF}_2:\text{Nd}^{3+}$ [J]. *Phys. Rev. B*, 1993, 47(22): 14706-14723.
- [25] ORLOVSKII Y V, FEDOROV V V, BASIEV T T, *et al.* Nonradiative relaxation and inhomogeneous splitting of aggregated optical centers in the Nd^{3+} -doped CaF_2 and SrF_2 crystals (FLN and decay study) [J]. *J. Lumin.*, 1999, 83-84: 361-366.
- [26] RUMELHART D E, HINTON G E, WILLIAMS R J. Learning representations by back-propagating errors [J]. *Nature*, 1986, 323(6088): 533-536.
- [27] HECHT-NIELSEN R. Theory of the backpropagation neural network [M]. WECHSLER H. *Neural Networks for Perception*. Boston: Academic Press, 1992: 65-93.
- [28] CHUN Y. A new axiomatization of the Shapley value [J]. *Games Econ. Behav.*, 1989, 1(2): 119-130.
- [29] BRATSCH S G, LAGOWSKI J J. Madelung constants, bond lengths, and coordination [J]. *J. Phys. Chem.*, 1985, 89(9): 1692-1695.
- [30] WANG C L, XU S H, YE L H, *et al.* The influence of solvent and ligands on characters of ZnS clusters [J]. *Struct. Chem.*, 2010, 21(6): 1215-1219.
- [31] WILDNER M. On the geometry of Co(II)O_6 polyhedra in inorganic compounds [J]. *Z. Kristallogr.*, 1992, 202(1-2): 51-70.
- [32] MURPHY L R, MEEK T L, ALLRED A L, *et al.* Evaluation and test of Pauling's electronegativity scale [J]. *J. Phys. Chem. A*, 2000, 104(24): 5867-5871.
- [33] BARANDIARÁN Z, SEIJO L. Quantum chemical analysis of the bond lengths in f^n and $f^{n-1}d^1$ states of Ce^{3+} , Pr^{3+} , Pa^{4+} , and U^{4+} defects in chloride hosts [J]. *J. Chem. Phys.*, 2003, 119(7): 3785-3790.
- [34] TAN S T, WANG R K, SONG G K, *et al.* Machine learning and Shapley Additive Explanation-based interpretable prediction of the electrocatalytic performance of N-doped carbon materials [J]. *Fuel*, 2024, 355: 129469.
- [35] SONG H L, LI Y T, ZOU X F, *et al.* Elite male table tennis matches diagnosis using SHAP and a hybrid LSTM-BPNN algorithm [J]. *Sci. Rep.*, 2023, 13(1): 11533.
- [36] AMIN M N, AHMAD W, KHAN K, *et al.* Evaluating the relevance of eggshell and glass powder for cement-based materials using machine learning and SHapley Additive exPlanations (SHAP) analysis [J]. *Case Stud. Constr. Mater.*, 2023, 19: e02278.
- [37] ZHANG S J, LEI H G, ZHOU Z C, *et al.* Fatigue life analysis of high-strength bolts based on machine learning method and SHapley Additive exPlanations (SHAP) approach [J]. *Structures*, 2023, 51: 275-287.



陈炳荣(1999-),男,广东清远人,硕士研究生,2022年于吉林大学珠海学院获得学士学位,主要从事固体发光材料的研究。

E-mail: 1759525074@qq.com



马凤凯(1988-),男,新疆阿克苏人,博士,副教授,博士生导师,2018年于中国科学院上海硅酸盐研究所获得博士学位,主要从事激光晶体材料的局域结构设计 with 性能调控以及人工智能驱动的激光晶体材料发展新范式探索。

Email: mafengkai@jnu.edu.cn



贺凤婷(1999-),女,广东梅州人,硕士研究生,2023年于广东技术师范大学获得学士学位,主要从事激光晶体材料的机器学习预测研究。

E-mail: 15766114737@163.com



苏良碧(1979-),男,湖北监利人,博士,研究员,博士生导师,2005年于中国科学院上海光学精密机械研究所获得博士学位,主要从事激光晶体材料的局域结构设计、晶体生长和激光应用研究。

Email: suliangbi@mail.sic.ac.cn